Super resolution and denoising with SRcycGAN

1 ABSTRACT

In recent years, the demand for high-quality images has surged across various fields, including medical imaging, satellite imagery, and consumer electronics. This study addresses the challenge of super-resolution (SR)—enhancing the resolution of low-quality images—using a Generative Adversarial Network (GAN)-based approach. Specifically, we employed the SRcycGAN model, known for its ability to generate realistic high-resolution images from low-resolution inputs. We conducted experiments using two datasets: DIV2K and the USA License Plate dataset, focusing on how pretraining and dataset characteristics influence model performance.

Our research highlights several key findings: First, pretraining on the DIV2K dataset significantly improved the model's performance, enhancing its generalization to new domains. Second, dataset size emerged as a critical factor, with larger training sets leading to more robust and accurate results. Third, while the model excelled in enhancing resolution, it struggled with noise reduction, particularly when dealing with complex distortions. The study concludes that a combination of pretraining on high-quality datasets and using diverse, large-scale data is essential for optimizing GAN-based models in super-resolution tasks. These insights provide a foundation for future work aimed at developing more versatile and effective models for real-world applications.

2 Introduction

In recent years, the demand for high-quality images has increased significantly across various fields, including medical imaging, satellite imagery, and consumer electronics. One of the key challenges in these areas is the problem of **super-resolution (SR)**—the task of enhancing the resolution of low-quality images to make them clearer and more detailed. This study focuses on addressing this problem by employing a Generative Adversarial Network (GAN) model, which has shown remarkable potential in producing high-quality, realistic images.

Generative Adversarial Networks (GANs) are a class of machine learning frameworks designed to generate new data samples that are indistinguishable from real data. A GAN consists of two main components: a generator and a discriminator. The generator creates images from noise, attempting to mimic real images, while the discriminator evaluates these images, distinguishing between real and generated ones. Through a process of adversarial training, where the generator and discriminator are pitted against each other, the generator gradually improves, producing increasingly realistic images. This unique capability of GANs makes them particularly well-suited for image-related tasks, including super-resolution.

The research question guiding this study is: **Can GAN models effectively enhance low-resolution images to achieve high-quality super-resolution?** This question is of critical importance in a world where visual clarity can impact decision-making in areas such as medical diagnosis, remote sensing, and media production. Traditional methods for super-resolution often rely on interpolation or other deterministic approaches that can result in blurring or loss of detail. In contrast, GANs have the ability to learn intricate features and generate high-resolution images that maintain the texture and sharpness of the original, low-resolution input.

Using GANs to solve the super-resolution problem is essential because of their unparalleled ability to generate photorealistic images with fine details. Unlike conventional techniques, GANs can produce results that are closer to what a human observer would perceive as natural and detailed. This capability is particularly important when working with images where preserving fine details is crucial, such as in medical imaging or satellite photos, where clarity can significantly influence outcomes. Therefore, employing GANs for super-resolution not only advances the field of image enhancement but also opens new possibilities for applications where high-resolution imagery is a necessity.

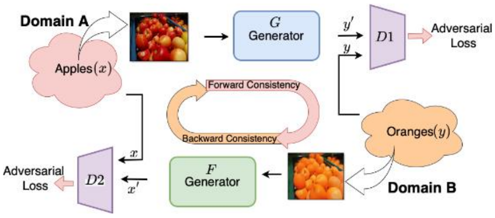

Figure 1: Cycle GAN architechture, from [3]

3 Literature Review

The development of our project is grounded in the exploration and extension of techniques presented in three key articles: **[1]** , [2] and [3]. These foundational works provide a robust framework for addressing the challenges of image super-resolution using advanced deep learning methods, particularly focusing on the application of GANs with residual and convolutional architectures.

3.1 The first article **"Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks"**[1] , introduces a novel approach to image translation tasks where paired training data is not available. This work extends the application of GANs to scenarios where aligning input and output images is challenging or impossible. The key innovation in this study is the cycle-consistency loss, which ensures that the translation from one domain to another and back again produces the original image, preserving essential content and structure. By leveraging adversarial training, the model generates more realistic and contextually appropriate images, even when faced with diverse and unpaired datasets. The cycle-consistent GAN (CycleGAN) framework presented in this paper has proven to be highly effective in a wide range of tasks, including style transfer, domain adaptation, and other forms of image manipulation, making it a versatile and robust tool for image-to-image translation across different domains.

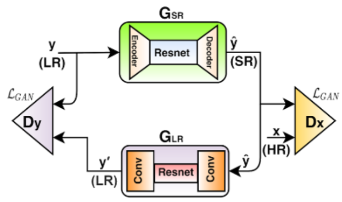

Figure 2: SRcycGAN architecture, from [2]

3.1 The second article, **"Deep Cyclic Generative Adversarial Residual Convolutional Networks for Real Image Super-Resolution,"** [2]introduces a novel approach, ***SRResCycGAN***, that leverages the strengths of cyclic GANs combined with residual convolutional networks. The cyclic structure allows for a more consistent and stable training process by incorporating a cycle-consistency loss, which helps the model maintain the original content and structure of the input images during the super-resolution process. The addition of residual connections within the network enhances the model's ability to learn complex features. This combination results in a more efficient and accurate reconstruction of high-resolution images from low-resolution inputs, making it highly suitable for real-world applications where image clarity is paramount.

3.3 The third article, **"Deep Generative Adversarial Residual Convolutional Networks for Real-World Super-Resolution"** [3] extends the application of GANs to more challenging and diverse real-world scenarios. This study addresses the limitations of traditional super-resolution techniques by incorporating adversarial training, which allows the model to produce more realistic and visually appealing images. The residual convolutional layers within the network are crucial for capturing fine details and textures, which are often lost in low-resolution images. By focusing on real-world images, this approach acknowledges and handles the variability and noise that are inherent in practical scenarios, making the model more robust and applicable across different domains.

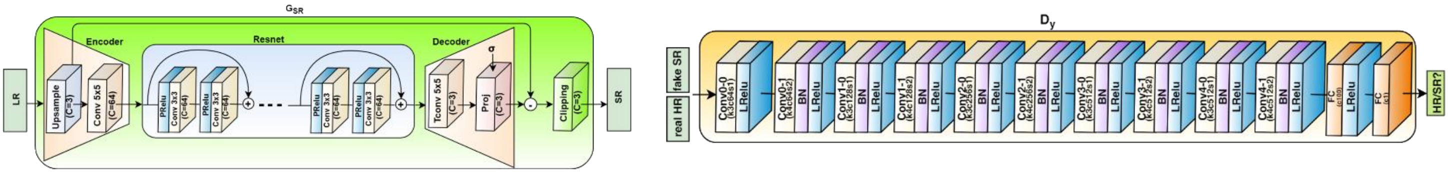

Figure3 : SRcycGAN generator and discriminator architecture, from: [2]

3.3 Integration and Extension

Our project builds on the methodologies and findings from these three articles, aiming to further refine the super-resolution process through the integration of advanced GAN architectures. We draw inspiration from the cyclic structure and residual networks discussed in these studies, while also exploring additional modifications to enhance performance in even more complex and diverse image datasets. The insights provided by these foundational works have been instrumental in shaping the direction of our research, particularly in terms of balancing the trade-offs between image quality, computational efficiency, and the ability to generalize across different types of images.

In conclusion, the literature provides a strong foundation for our project, highlighting the effectiveness of GAN-based approaches in super-resolution tasks. By building on the principles established in these studies, our work seeks to push the boundaries of what can be achieved in image enhancement, particularly in real-world applications where high-resolution imagery is essential.

4 Methods

4.1 Datasets

For this study, we employed two distinct datasets to evaluate the effectiveness of our super-resolution models: the **DIV2K dataset** and the **USA License Plate dataset**.

4.1.1 **DIV2K Dataset**: The DIV2K dataset is a widely recognized benchmark for image super-resolution tasks, consisting of 800 high-resolution images for training and 100 images each for validation and testing. The images cover a broad range of scenes and textures, making it ideal for training models that need to generalize well to diverse image types.

Figure 4: DIV2K Dataset

4.1.2 **USA License Plate Dataset**: The USA License Plate dataset was used to test the model's performance on a more specific and challenging domain, where clarity of small text and fine details is crucial. This dataset includes images of vehicle license plates taken under various conditions, such as different lighting, angles, and resolutions.

Figure 5: USA Lisence Plates Dataset

4.2 Data Preprocessing

4.2.1 Downsampling:

Images were downscaled by a factor of 4 using the BICUBIC DISTORTION algorithm, which preserves more image details, essential for super-resolution tasks.

4.2.2 Dimensional Distortion:

To simulate real-world spatial distortions, we applied random rotations (30-70 degrees) around the Y-axis, generating multiple distorted versions of each image to enhance the model's robustness.

Preprocessing for article's model:

4.4.3 Noisy Image Creation:

Noisy images were generated using DSGAN, a network designed to estimate and reduce noise, aiding the model in learning to handle noise effectively.

4.4.4 Data Repetition:

The dataset was tripled by processing each image three times per batch, stabilizing training with smoother gradients. While this increases dataset size and feature learning, it also risks overfitting by reinforcing repetitive patterns.

4.4.5 Data Mixup:

Images in each batch were mixed with others using a random proportion from a beta distribution. This technique enhances generalization and robustness but requires careful tuning to avoid underfitting and blurring.

4.4.6 Image Chopping:

To manage GPU memory, images were recursively chopped into smaller patches (minimum 100,000 pixels) before processing. This approach prevents memory errors but slows down the testing process.

4.3 Model Training with Various Hyperparameters

4.3.1 Learning Rate:

We fine-tuned the learning rate to balance training stability and convergence, using lower rates for more precise adjustments.

4.3.2 Batch Size:

Batch size was adjusted to balance between stable gradient updates and memory usage, with larger sizes providing smoother learning.

4.3.3 Optimizer:

Adam was selected for its adaptive learning rate, offering efficient and stable convergence throughout training.

4.3.4 Number of Epochs:

Epochs were carefully chosen to avoid underfitting or overfitting, ensuring the model learned effectively while generalizing well to new data.

4.4 Model Development

We adopted a dual approach in our methodology, leveraging both an existing model from the literature and a custom-built model:

4.4.1 Existing Model: To assess and compare our model's performance against a state-of-the-art model, we conducted experiments with both the SRResCycGAN model and the model described in [3]. For the SRResCycGAN model, we utilized the original code provided by Rao Muhammad

Umer et al. and trained it on 500 images from the USA License Plates dataset. The training images were augmented, tripling the dataset to 1500 images. These were divided into batches of 16, resulting in 93 batches per epoch. The training process was set to run for 10,000 iterations, which equated to approximately 107 epochs. Throughout the training, the learning rate for both the generator and discriminator was halved at specific intervals to refine the model's performance. In addition, we implemented the model from [3] , following the architecture and training protocols detailed in the paper. This model uses residual convolutional layers combined with adversarial training techniques to enhance the resolution of low-quality images. Our implementation significantly mirrored the original work to ensure a fair comparison between the models.

4.4.2 **Custom-Built Model**: In addition to using the existing model, we developed a new GAN-based model from scratch. Our model design was inspired by the cyclic GAN and residual network principles discussed in the article **"Deep Cyclic Generative Adversarial Residual Convolutional Networks for Real Image Super-Resolution."** However, we introduced several modifications aimed at improving the model's performance on more complex and varied datasets. These modifications included optimizing the generator's architecture for better feature extraction and implementing advanced techniques for stabilizing the adversarial training process.

4.5 Training and Evaluation

Both models were trained on the **DIV2K dataset** using the same set of hyperparameters to ensure a fair comparison. The models were trained for a predetermined number of epochs, with the performance being evaluated on a separate validation set after each epoch.

After initial training on the DIV2K dataset, we fine-tuned both models on the **USA License Plate dataset** to assess their ability to generalize to a different domain. This fine-tuning involved further training with a smaller learning rate to refine the models' performance on the new dataset.

To comprehensively evaluate model performance, we tested 100 images and computed an average for each metric, enabling a direct comparison to the baseline performance reported in the referenced article.

Train	Dataset	Number of images	Epochs	Generator Learning Rate	Discriminator Learning Rate	Loaded Train Model No. Weights	Fine Tuning
9	Car Plates	750	30	$2x10^{-4}$	$1x^{-}10^{5}$	X	X
10	DIV2K	800	30	$2x10^{-4}$	$1x^{-}10^{5}$	9	X
11	Car Plates	750	30	Changing learning rate starting with $1x^{-}10^{5}$	X	10	V

Table 1: Final trainings (no distortion)

Train	Dataset	Number of images	Epochs	Generator Learning Rate	Discriminator Learning Rate	Loaded Train Model No. Weights	Fine Tuning
12	Car Plates Distorted	750	30	Changing learning rate starting with 1x-105	X	9	V
13	Car Plates Distorted	750	30		X	10	V
14	Car Plates Distorted (HR)	750	30	2x10-4	1x-105	X	X

Table ***2: Final trainings (with distortion)***

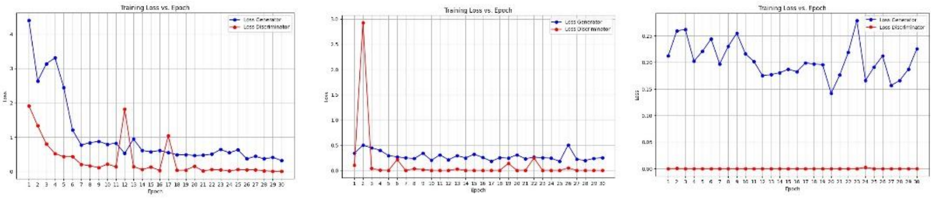

Figure 6: Generatoe VS discriminator Loss graphs

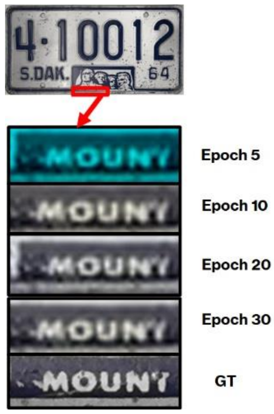

Figure ***7: model progression***

4.4 Metrics

To evaluate the performance of the models, we used several widely accepted metrics for image super-resolution tasks:

4.4.1 Peak Signal-to-Noise Ratio (PSNR): To measure the quality of the reconstructed high-resolution images in comparison to the ground truth.
4.4.2 Structural Similarity Index (SSIM): To assess the perceptual similarity between the original and the super-resolved images.
4.4.3 Learned Perceptual Image Patch Similarity (LPIPS): To evaluate the perceptual quality of the images as perceived by human observers.

By comparing these metrics across different epochs and datasets, we aimed to determine the strengths and weaknesses of each model in both general and specific image enhancement tasks.

4.5 Results

Experiment	Distortion	Model	PSNR↑	SSIM↑	LPIPS↓	Result Output
1	X	Our Model	38	0.985	0.013	
2	X	Our Model	40.4	0.992	0.0072	
3	V	Our Model	11.68	0.348	0.657	
4	V	Our Model	24	0.88	0.119	
5	X	Article Model	25.95	0.735	0.255	

*Table 3: **Test results, best model in red square***

***Figure 8**: The “PUB” part compared. Top: The Ground-Truth high-resolution image, where the PUB part is zoomed in X68. Middle: The super resolution image (output of the model) where the “PUB” part zoomed ’68. Bottom: The low-resolution image, ere the “PUB” part zoomed in X313.*

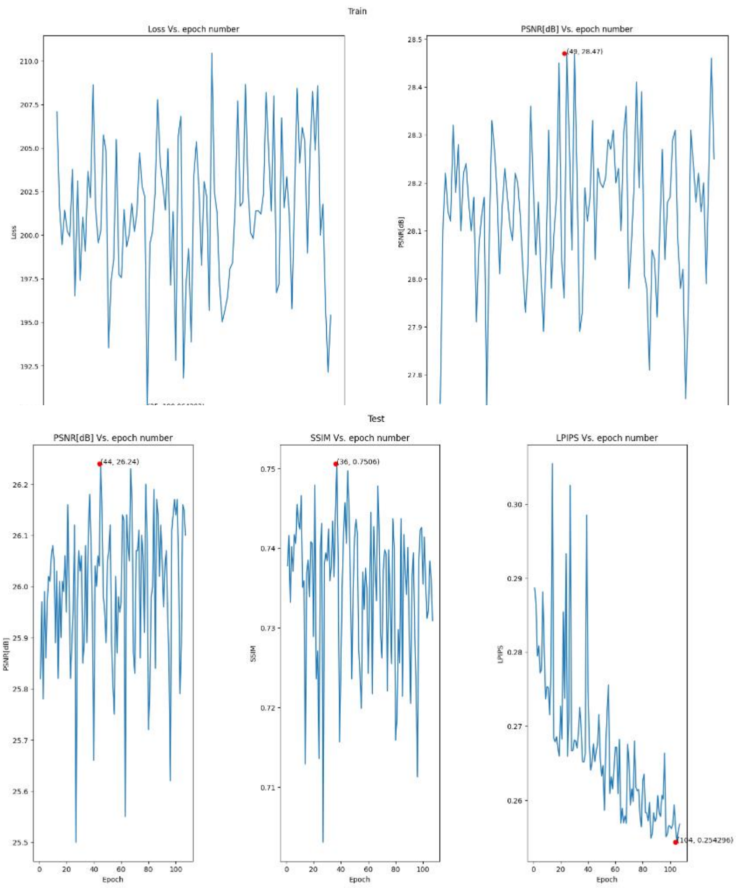

Figure 9: Loss and metrics depend on epoch [article's model]

5 Discussions

5.1 Model Flexibility: The performance of the SRcycGAN model in our study demonstrated notable variability depending on the type of image resolution it was trained on. Specifically, the model exhibited strong results when addressing bicubic distortion, a common type of downscaling used in super-resolution tasks. However, its effectiveness was significantly reduced when dealing with dimensional distortions, where the aspect ratio or other geometric properties of the image were altered. This suggests that while SRcycGAN is well-suited for certain standard types of image degradation, its ability to generalize to more complex or varied distortions remains limited. This highlights the need for further refinement in the model architecture or training procedures to enhance its flexibility and robustness across different types of distortions.

5.2 Generalization: Pretraining the SRcycGAN model on the DIV2K dataset before fine-tuning on the USA License Plate dataset proved beneficial, as it allowed the model to start with a stronger foundation and adapt more quickly to the specific characteristics of the new data. However, despite the improvements observed through pretraining, the model's ability to generalize effectively to new datasets still heavily depends on the similarity of the training data to the new data. When the dimensions or characteristics of the new data significantly differ from the original training set, the model's performance tends to decline. This dependency suggests that while pretraining is a valuable technique, the diversity and representativeness of the training data are crucial for achieving robust generalization in real-world applications.

5.3 Article's model performance: We can see that the loss function as well as the PSNR suffer from fluctuations and is very unstable, not presenting any meaningful trend of changing, so it seems like the model suffers from inefficient learning. It could be explained by the fact that the model is pretrained and training it even more didn’t make a significant change, or it could be a result of the data repetition which makes the model memorize patterns.

5.3 Future Work: To further enhance the capabilities of the SRcycGAN model, several avenues for future research and development are proposed:

- **5.3.1** **Test with More Datasets**: A key area for future exploration involves expanding the range of datasets used to train and evaluate the model. By incorporating a wider variety of datasets, encompassing different types of images and resolutions, we can better understand how well the model generalizes across a larger spectrum of scenarios. This could lead to the development of a more versatile model capable of handling a broader array of super-resolution tasks.
- **5.3.2** **Add More Distortions**: Another critical direction for future work is to introduce more complex and varied distortions during the training process. By exposing the model to a greater diversity of distortions, including those that more accurately reflect real-world conditions, we can potentially improve its ability to handle the unpredictable and often challenging distortions encountered in practical applications. This could involve simulating conditions such as motion blur, noise, or lens distortion, thereby equipping the model to deliver higher quality super-resolution results in a wider range of real-world scenarios.

By pursuing these future research directions, we aim to build upon the foundational work of SRcycGAN, ultimately creating a more robust and adaptable model for image super-resolution.

6 Conclusions

Our study yielded several key insights into the factors that influence the performance of the SRcycGAN model in the task of image super-resolution:

6.1 Pretraining on DIV2K: Pretraining the SRcycGAN model on the DIV2K dataset significantly improved its performance, leading to better generalization and higher quality outputs. This foundational training allowed the model to develop a strong baseline, which was crucial for adapting to new data with similar characteristics.

6.2 Dataset Size: The size of the training dataset proved to be a critical factor in determining the effectiveness of the model. Larger training datasets contributed to superior model results, enhancing the model's robustness and the fidelity of the reconstructed images.
This underscores the importance of ample and diverse data when training GANs for complex tasks like super-resolution.

6.3 Distortion Handling: While training with various distortions made the model less effective at denoising, it remained highly capable of reconstructing high-resolution images from low-resolution inputs. This indicates that while the SRcycGAN model excels at enhancing resolution, additional strategies may be needed to improve its ability to handle noise and other specific types of distortion.

6.4 Optimal Configuration: The best overall performance in the SRcycGAN model was achieved by combining pretraining on a high-quality dataset like DIV2K with the use of a large and diverse training dataset. This optimal configuration allowed the model to deliver the most consistent and high-quality results, demonstrating the importance of both pretraining and dataset size in the development of effective super-resolution models.

These findings provide valuable guidance for future work in the field, highlighting the critical role of data quality, quantity, and training strategies in optimizing GAN-based models for image super-resolution.

7 References

[1] Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks*. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) (pp. 2223-2232). Berkeley AI Research (BAIR) Laboratory, UC Berkeley.

[2] Umer, R. M., & Micheloni, C. (2020). *Deep Cyclic Generative Adversarial Residual Convolutional Networks for Real Image Super-Resolution*. University of Udine, Italy.

[3] Umer, R. M., & Micheloni, C. (2020). *Deep Generative Adversarial Residual Convolutional Networks for Real-World Super-Resolution*. University of Udine, Italy.